The Gallagher Committee and at least one of the witnesses seemed unwilling to accept the distinction between the short-run and the long-run on the one hand, and the distinction between the statistical information systems and intelligence information systems on the other.

This skepticism grows out of (1) the fact that a statistical system must contain information about individual respondents, thereby rendering it potentially useful for intelligence purposes; (2) the fact that no system designed exists for providing *foolproof* protection against file misuse in this way; and (3) the argument that the pace of technology is proceeding so fast that there will be no technical limitations on accumulating "all of the data about everyone." They would claim, therefore, that the long run is already upon us.

The author would still maintain that these distinctions are valid and vital. It is true that with respect to *intelligence* systems, there is no point in making any effective distinction between the short-run and the long-run. The detailed and careful standards with respect to collecting, coding, and tabulating data essential to a meaningful matching of attributes for statistical analysis is not essential for single-record searches for intelligence purposes. Therefore, the development of intelligence systems is not constrained by a prior requirement for production reforms.

Furthermore, the system design, from both the hardware and software point of view, essential to perform a satisfactory intelligence retrieval function is already available for some quite large files and capable of rapid development. As Mr. Baran (a computer expert from RAND Corporation) pointed out the Gallagher Committee, we are already building the bits and pieces of an intelligence data bank that can be quite easily merged. He pointed out that no one planned a national railway system. It started out as short routes connecting local population centers. It gradually merged into a larger system. In effect, an intelligence system may be 50 per cent in being already. In substance, he is saying "instead of distinguishing between a short-run and long-run it is essential to realize that it may be later than you think." Mr. Baran is right with respect to intelligence systems. He is wrong, we think, with respect to the threat of misuse of a statistical system in the short-run. Consider briefly the reasons:

(1) As we have seen, a statistical information system of greater utility for policy cannot be developed without making substantial changes in production practices. This yields the necessity for changes to come slowly. Subsets of traditional files will need to be modified and integrated for matching purposes on an incremental basis with priorities established by important requirements.

Mr. Baran's railway or telegraph merge concept does not apply so simply here. We might transform the analogy to provide a more legitimate comparision if we visualize the development of the separate railway systems having taken place in such a way that each segment was technically incompatible with each other segment so that the system of lines might require something approximating a 100% replacement capital through obsolescence if the system were to be integrated. The present problem the world faces in integrating incompatible TV systems or incompatible systems of measurement might form a closer analogy. In short, no one is going to plan a complete integration of all statistical records over any short-run period of time. It would cost a great deal, and regardless of cost it would take a considerable period of time to put into effect.

(2) An *intelligence system*, if it is going to be efficient, has to be as nearly complete as possible. Ideally it should constitute a census so that every possible individual search request could be fulfilled. There has been an implicit assumption in committee hearings that this is also true of a statistical system, but it emphatically is not. We have found increasingly that the efficient statistical system (since it generates related information about groups of people and never about individuals) doesn't want "all the data on everyone." It only wants some of the data on some of the people—enough to be relevant for the important problems of analysis by private business, government, and researchers and enough to support reliable inferences. To build a complete file is inordinately expensive, and we have found, for most statistical purposes less reliable. Indeed, the national "census," since it is conductd primarily for statistical purposes rather than intelligence purposes, is a complete census for only a very few attributes of the population. The bulk of information is collected on a sample basis only.

(3) The existing statistical systems have had considerable experience and an admirable record in protecting personal privacy through legal regulations supplemented by operational procedures. Initial moves to improve the matching characteristics of federal statistical records *for statistical purposes* could be carried out under an extension of well-established protection procedures.